

Modelos Lineares Generalizados Duplos

Gustavo Almeida e Pedro Almeida

Universidade Federal de Juiz de Fora

14/12/2023

- Modelos Gaussianos
- Modelos Lineares Generalizados
- MLG Duplo - Modelagem de Locação e Escala

Família Exponencial

$$f(y; \theta_i, \Phi) = \exp[\Phi y \theta_i - b(\theta_i) + c(y, \Phi)]$$

Preditor Linear de μ

$$\eta_i = \mathbf{X}_i^T \boldsymbol{\beta}$$

Preditor Linear de ϕ

$$\lambda_i = \mathbf{Z}_i^T \boldsymbol{\gamma}$$

Além disso, cada preditor linear possui uma função de ligação

Função de Ligação de μ

$$g(\mu_i) = \eta_i$$

Função de Ligação de ϕ

$$h(\phi) = \lambda(X_i)$$

- Método da Máxima Verossimilhança (Smyth 1989)
- Método da Máxima Verossimilhança Restrita (Smyth and Verbyla 1999) (McCullagh and Tibshirani 1990)

Método via Máxima Verossimilhança

Como apresentado em (Paula 2013), a função escore e a matriz informação de Fisher para β podem ser obtidas calculando as derivadas da função de verossimilhança, da seguinte forma

$$\delta L(\boldsymbol{\theta})/\delta\beta_j = \sum_{i=1}^n \left\{ \frac{\omega_i}{V_i} (y_i - \mu_i) x_{ij} \right\}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2$. Dessa forma, temos a função escore dada por

$$U_{\beta}(\boldsymbol{\beta}) = \frac{\delta L(\boldsymbol{\theta})}{\delta\boldsymbol{\beta}} = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$$

em que \mathbf{X} é uma matriz $n \times p$ de posto completo, $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ é a matriz de pesos,

$$\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}, \mathbf{y} = \{y_1, \dots, y_n\} e \boldsymbol{\mu} = \{\mu_1, \dots, \mu_n\}.$$

$$\begin{aligned} \partial^2 L(\theta)/\partial\beta_j\partial\beta_\ell &= \phi \sum_{i=1}^n (y_i - \mu_i) \frac{d^2\theta_i}{d\mu_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij}x_{i\ell} \\ &+ \phi \sum_{i=1}^n (y_i - \mu_i) \frac{d\theta_i}{d\mu_i} \frac{d^2\mu_i}{d\eta_i^2} x_{ij}x_{i\ell} - \phi \sum_{i=1}^n \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij}x_{i\ell} \end{aligned}$$

dessa forma, obtemos o valor esperado

$$\begin{aligned} E \{ \partial^2 L(\theta)/\partial\beta_j\partial\beta_\ell \} &= -\phi \sum_{i=1}^n \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij}x_{i\ell} \\ &= -\phi \sum_{i=1}^n \frac{(d\mu_i/d\eta_i)^2}{V_i} x_{ij}x_{i\ell} \\ &= -\phi \sum_{i=1}^n \omega_i x_{ij}x_{i\ell}. \end{aligned}$$

Escrevendo na forma matricial, concluímos

$$\mathbf{K}_{\beta\beta}(\theta) = \mathbf{E} \left\{ -\frac{\partial^2 L(\theta)}{\partial\beta\partial\beta^T} \right\} = \phi\mathbf{X}^T\mathbf{W}\mathbf{X}$$

Agora, deve-se realizar o mesmo procedimento na obtenção dos estimadores de máxima verossimilhança para γ , em que

$$\begin{aligned} \partial L(\theta) / \partial \gamma_j &= \sum_{i=1}^n \left\{ \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_j} t_i + d'(\phi_i) \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{1}{h'(\phi_i)} z_{ij} t_i + d'(\phi_i) \frac{1}{h'(\phi_i)} z_{ij} \right\} \\ &= \sum_{i=1}^n \frac{z_{ij}}{h'(\phi_i)} \{t_i + d'(\phi_i)\} \end{aligned}$$

em que $h'(\phi_i) = d\lambda_i/d\phi_i$. Portanto, em forma matricial obtemos

$$\mathbf{U}_\gamma = \mathbf{Z}^T \mathbf{H}_\gamma^{-1} (\mathbf{t} - \mu_T),$$

em que $\mathbf{H}_\gamma = \text{diag} \{h'(\phi_1), \dots, h'(\phi_n)\}$, $\mathbf{t} = (t_1, \dots, t_n)^T$ e $\mu_T = (\mathbb{E}(T_1)) = (-d'(\phi_1), \dots, -d'(\phi_n))^T$.

Na obtenção da matriz informação de Fisher para γ , vamos calcular as derivadas segundas da função log verossimilhança

$$\begin{aligned} \partial^2 L(\theta) / \partial \gamma_j \partial \gamma_\ell &= - \sum_{i=1}^n \frac{z_{ij}}{\{h'(\phi_i)\}^2} \left[d''(\phi_i) h(\phi_i) \frac{d\phi_i}{d\lambda_i} z_{il} - h''(\phi_i) \{t_i + \dots\} \right] \\ &= - \sum_{i=1}^n \frac{z_{ij} z_{il}}{\{h'(\phi_i)\}^2} \left[d''(\phi_i) - \frac{h''(\phi_i)}{h'(\phi_i)} \{t_i + d'(\phi_i)\} \right] \end{aligned}$$

e cujos valores esperados ficam dados por

$$E \left\{ - \frac{\partial^2 L(\theta)}{\partial \gamma_j \partial \gamma_\ell} \right\} = - \sum_{i=1}^n \frac{d''(\phi_i)}{\{h'(\phi_i)\}^2} z_{ij} z_{il}.$$

Em forma matricial obtemos

$$\mathbf{K}_{\gamma\gamma} = \mathbf{Z}^T \mathbf{P} \mathbf{Z}$$

em que $\mathbf{P} = \mathbf{V}_\gamma \mathbf{H}_\gamma^{-2}$, $\mathbf{V}_\gamma = \text{diag} \{-d''(\phi_1), \dots, -d''(\phi_n)\}$.

Devido à ortogonalidade entre os parâmetros θ_i e ϕ_i , segue diretamente a ortogonalidade entre β e γ . Assim, a matriz de informação de Fisher para θ é dado por $\mathbf{K}_{\theta\theta} = \text{diag}\{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\gamma\gamma}\}$.

Assim como na teoria dos modelos lineares generalizados, podemos estabelecer um método iterativo para encontrar $\hat{\beta}$ e $\hat{\gamma}$. De acordo com [?], é possível atualizar essas estimativas a partir das expressões

$$\beta^{(m+1)} = \left(\mathbf{X}^T \Phi^{(m)} \mathbf{W}^{(m)} \mathbf{X} \right)^{-1} \mathbf{X}^T \Phi^{(m)} \mathbf{W}^{(m)} \mathbf{y}^{*(m)}$$

$$\gamma^{(m+1)} = \left(\mathbf{Z}^T \mathbf{P}^{(m)} \mathbf{Z} \right)^{-1} \mathbf{Z}^T \mathbf{P}^{(m)} \mathbf{z}^{*(m)}$$

em que $\mathbf{y}^* = \mathbf{X}\beta + \mathbf{W}^{-1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \mu)$, $\mathbf{z}^* = \mathbf{Z}\gamma + \mathbf{V}_\gamma^{-1} \mathbf{H}_\gamma (\mathbf{t} - \mu_T)$ e $m = 0, 1, 2, \dots$

Conforme [?] esse processo iterativo pode ser resolvido alternando-se as duas equações até a convergência. No chute inicial para ϕ_i pode-se considerar $\phi_i = \hat{\phi}, \forall i = 1, \dots, n$, onde $\hat{\phi}$ é a estimativa de ϕ obtida de um MLG.

Sob as condições de regularidade, temos que para n grande $\hat{\beta} \sim N_p(\beta, \mathbf{K}_{\beta\beta}^{-1})$ e $\hat{\gamma} \sim N_q(\gamma, \mathbf{K}_{\gamma\gamma}^{-1})$, respectivamente. Além disso, devido à ortogonalidade entre β e γ segue a independência assintótica entre $\hat{\beta}$ e $\hat{\gamma}$.

Método via Máxima Verossimilhança Restrita

- É um método menos custoso
- Atualmente é utilizado nos pacotes DGLM e GAMLSS do R
- Utilizada uma abordagem

O estimador de MQO de γ é

$$Z^T W_d Z \gamma = Z^T W_d \mathbf{z}_d$$

Onde

$$\mathbf{W}_d = \text{diag}\left(\frac{1}{h(\phi_i)^2 V_d(\phi_i)}\right)$$

Via expansão de Taylor $var(\hat{\gamma})$ é dada por (Smyth and Verbyla 1999) (McCullagh and Tibshirani 1990)

$$var(\hat{\gamma}) \approx I_{\gamma\gamma}^{-1}$$

Com

$$I_{\gamma\gamma} = \frac{1}{2} \mathbf{Z}^T \mathbf{W}_d^* \mathbf{Z}$$

E

$$\mathbf{W}_d^* = \mathbf{W}_d - 2diag\left(\frac{h_i}{\phi^2 h(\phi_i)_i^2}\right) + H^2$$

- É um método direto de conversão da estimação via máxima verossimilhança de γ para uma aproximação via máxima verossimilhança restrita.
- Além disso, o método nos dá que $\hat{\gamma}$ seja aproximadamente não viciado, e com que o submodelo de dispersão nos dê os erros padrão ajustados.

- Dado a complexidade de estimação de H^2 , (Smyth and Verbyla 1999) recomenda uma aproximação via $diag(h_i^2)$
- Assim, a nova matriz de pesos é dada por

$$W_d^* \approx W_d - 2diag\left(\frac{h_i}{\phi_i^2 h(\phi_i)^2} + h_i^2\right)$$

Os desvios para a média são dados por

$$D_1^*(\mathbf{y}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\phi}}) = \sum_{i=1}^n d_1^{*2}(y_i; \hat{\mu}_i, \phi_i)$$

em que

$$d_1^{*2}(y_i; \hat{\mu}_i, \phi_i) = 2\phi_i \left[y_i (\tilde{\theta}_i - \hat{\theta}_i) + \{b(\hat{\theta}_i) - b(\tilde{\theta}_i)\} \right]$$

Para ϕ_i grande $\forall i$ o desvio $D_1^*(\mathbf{y}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\phi}})$ pode ser comparado com os percentis da distribuição X^2 com $(n - p)$ graus de liberdade.

O resíduo Studentizado, no caso dos MLGs duplos será dado por

$$t_{D_{1i}} = \frac{d_1^* (y_i; \hat{\mu}_i, \hat{\phi}_i)}{\sqrt{1 - \hat{h}_{ii}}},$$

Em que $d_1^* (y_i; \hat{\mu}_i, \hat{\phi}_i) = \pm \sqrt{d_1^{*2} (y_i; \hat{\mu}_i, \hat{\phi}_i)}$, o sinal continua sendo o mesmo de $(y_i - \hat{\mu}_i)$ e \hat{h}_{ii} é o i -ésimo elemento da diagonal principal da matriz.

Os desvios para a precisão são dados por

$$D_2^*(\mathbf{y}; \hat{\phi}, \mu) = \sum_{i=1}^n d_2^{*2} (y_i; \hat{\phi}_i, \mu_i)$$

Em que

$$d_2^{*2} (y_i; \hat{\phi}_i, \mu_i) = 2 \left[t_i (\tilde{\phi}_i - \hat{\phi}) + \{d(\tilde{\phi}_i) - d(\hat{\phi}_i)\} \right], \tilde{\phi}_i$$

é solução para ϕ_i sob o modelo saturado sendo dada por $d'(\tilde{\phi}_i) = -t_i$.

Para ϕ_i grande $\forall i$ o desvio $D_2^*(\mathbf{y}; \mu, \hat{\phi})$ pode ser comparado com os percentis da distribuição X^2 com $(n - p)$ graus de liberdade.

O resíduo componente do desvio para a precisão fica dado por

$$t_{D_2i} = \frac{d_2^* (y_i; \hat{\phi}_i, \hat{\mu}_i)}{\sqrt{1 - \hat{r}_{ii}}},$$

Em que $d_2^* (y_i; \hat{\phi}_i, \hat{\mu}_i) = \pm \sqrt{d_2^{*2} (y_i; \hat{\phi}_i, \hat{\mu}_i)}$, o sinal sendo o mesmo de $\{\hat{t}_i + d'(\hat{\phi}_i)\}$ e \hat{r}_{ii} é o i -ésimo elemento da diagonal principal da matriz

$$\hat{\mathbf{R}} = \hat{\mathbf{P}}^{1/2} \mathbf{Z} (\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\mathbf{P}}^{1/2},$$

A fim de analisar os resíduos, (Paula 2013) sugere o gráfico normal de probabilidades e contra os valores ajustados para t_{D_1} e t_{D_2} .

Em um MLG duplo, calcula-se LD_i , como nos modelos MLG, porém com $\hat{\phi}_i$ no lugar de $\hat{\phi}$, ou seja, esta será definida por

$$LD_i^\beta = \left\{ \frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \right\} t_{S_i}^2,$$

Em que

$$t_{S_i} = \frac{\sqrt{\hat{\phi}_i} (y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i (1 - \hat{h}_{ii})}}$$

Uma medida para avaliar a influência nas estimativas dos parâmetros da precisão fica dada por

$$\begin{aligned} LD_i^\gamma &= (\hat{\gamma}_{(i)} - \hat{\gamma})^T (\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z}) (\hat{\gamma}_{(i)} - \hat{\gamma}) \\ &= \left\{ \frac{\hat{r}_{ii}}{1 - \hat{r}_{ii}} \right\} t_{Ti}^2, \end{aligned}$$

Em que

$$t_{Ti} = \frac{t_i + d'(\hat{\phi}_i)}{\sqrt{-d''(\hat{\phi}_i)(1 - \hat{r}_{ii})}}.$$

A fim de analisar possíveis pontos de influência, (Paula 2013) sugere o gráfico normal de probabilidades e contra os valores ajustados para LD_i^β e LD_i^γ .

Os testes de hipóteses para um MLG Duplo são apenas uma extensão dos testes de hipóteses para um MLG padrão, considerando ϕ_i não constante o teste da razão de verossimilhanças para β dado por

$$\xi_{RV} = 2 \left\{ L(\hat{\beta}, \hat{\gamma}) - L(\beta^0, \hat{\gamma}) \right\}.$$

Da mesma forma, o teste em relação a γ é dado por

$$\xi_{RV}^{\gamma} = 2 \left\{ L(\hat{\beta}, \hat{\gamma}) - L(\hat{\beta}, \gamma^0) \right\}.$$

O teste de Wald para β é definido, nesse caso, por

$$\xi_W = [\hat{\beta} - \beta^0]^T \hat{\text{Var}}^{-1}(\hat{\beta}) [\hat{\beta} - \beta^0],$$

da mesma forma para γ

$$\xi_W^\gamma = [\hat{\gamma} - \gamma^0]^T \hat{\text{Var}}^{-1}(\hat{\gamma}) [\hat{\gamma} - \gamma^0],$$

- Foram utilizados 2 exemplos, distribuições Normal e Gama
- Os tamanhos amostrais: 30, 100, 500 e 1000 foram simulados 1000 vezes cada
- Os seguintes β e γ foram fixadas para ambas as distribuições

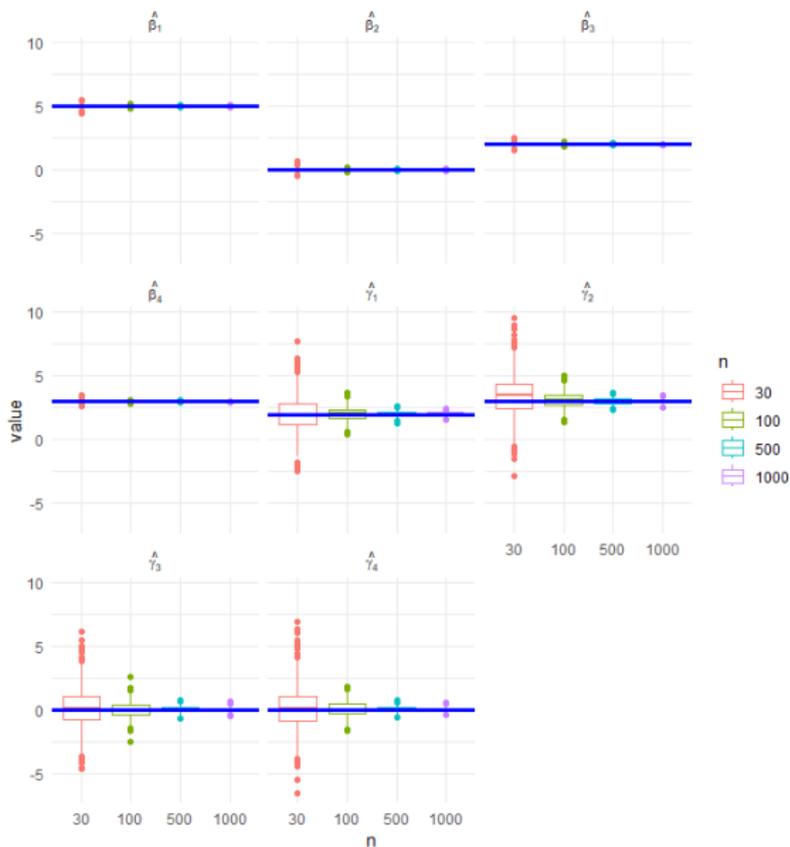
$$\beta^T = (5, 0, 2, 3)$$

$$\gamma^T = (2, 3, 0, 0)$$

O estudo de simulação se baseou no seguinte algoritmo:

- 1 Criação de um vetor de tamanhos amostrais e fixação dos valores de β e γ verdadeiros
- 2 Cálculo dos componentes lineares $\eta = X^T \beta$ e $\lambda = X^T \gamma$
- 3 Geração de valores de uma distribuição Normal ($\mu_i = \eta_i$, $\sigma_i^2 = \frac{1}{e^{\lambda_i}}$) e Gama ($\alpha_i = e^{\lambda_i}$, $\theta_i = \frac{e_i^{\lambda}}{e_i}$)
- 4 Estimação de β e γ via máxima verossimilhança
- 5 Construção de gráficos e cálculo de valores de acurácia de $\hat{\beta}$ e $\hat{\gamma}$

Modelo Normal



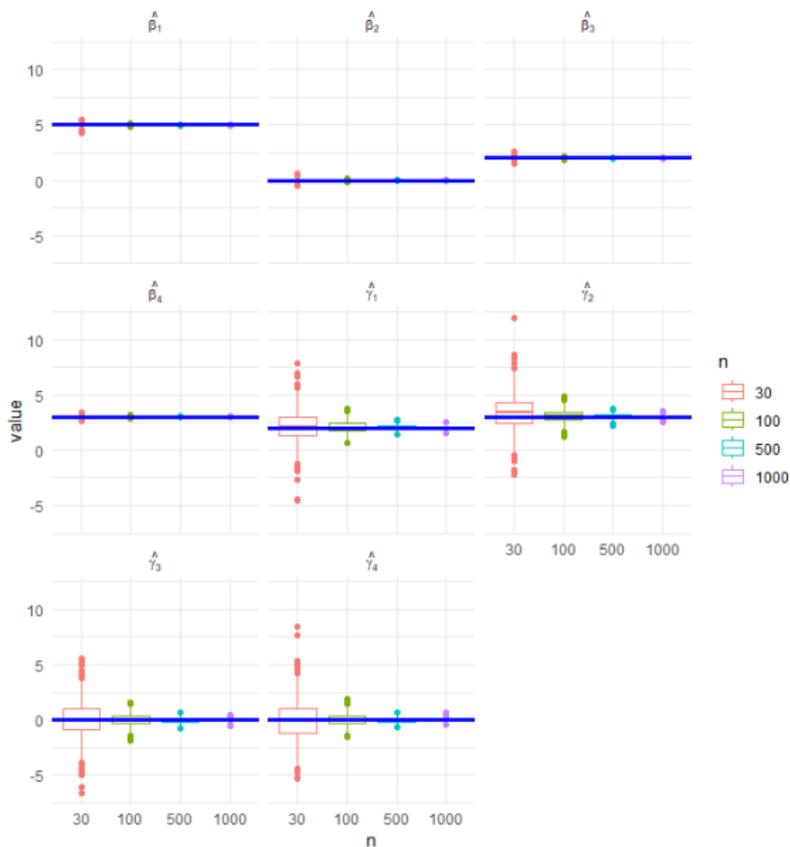
EQM do Modelo Normal

Parâmetro	n = 30	n = 100	n = 500	n = 1000
$\hat{\beta}_1$	0.02	0.00	0.00	0.00
$\hat{\beta}_2$	0.02	0.00	0.00	0.00
$\hat{\beta}_3$	0.01	0.00	0.00	0.00
$\hat{\beta}_4$	0.02	0.00	0.00	0.00
$\hat{\gamma}_1$	2.02	0.27	0.04	0.02
$\hat{\gamma}_2$	2.68	0.32	0.05	0.03
$\hat{\gamma}_3$	2.71	0.35	0.05	0.03
$\hat{\gamma}_4$	2.74	0.33	0.05	0.03

Considerações

- EQM de $\hat{\beta}$ se aproxima de 0 já para tamanhos amostrais baixos
- EQM de $\hat{\gamma}$ se aproxima de 0 para um n a partir de 500, evidenciando sua complexidade

Modelo Gama



EQM do Modelo Gama

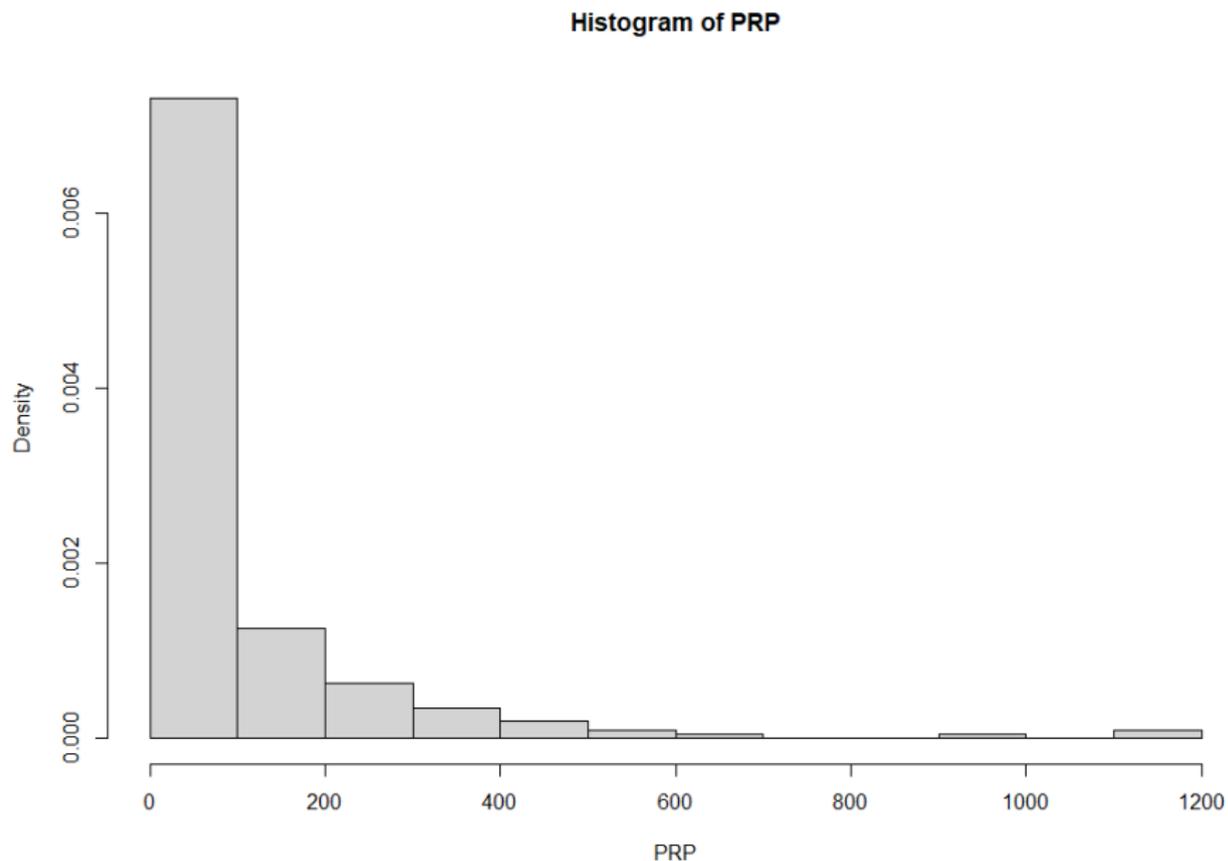
Parâmetro	n = 30	n = 100	n = 500	n = 1000
$\hat{\beta}_1$	0.02	0.00	0.00	0.00
$\hat{\beta}_2$	0.02	0.00	0.00	0.00
$\hat{\beta}_3$	0.01	0.00	0.00	0.00
$\hat{\beta}_4$	0.02	0.00	0.00	0.00
$\hat{\gamma}_1$	1.85	0.25	0.04	0.02
$\hat{\gamma}_2$	2.36	0.32	0.05	0.02
$\hat{\gamma}_3$	2.49	0.29	0.05	0.03
$\hat{\gamma}_4$	2.47	0.29	0.05	0.02

Considerações

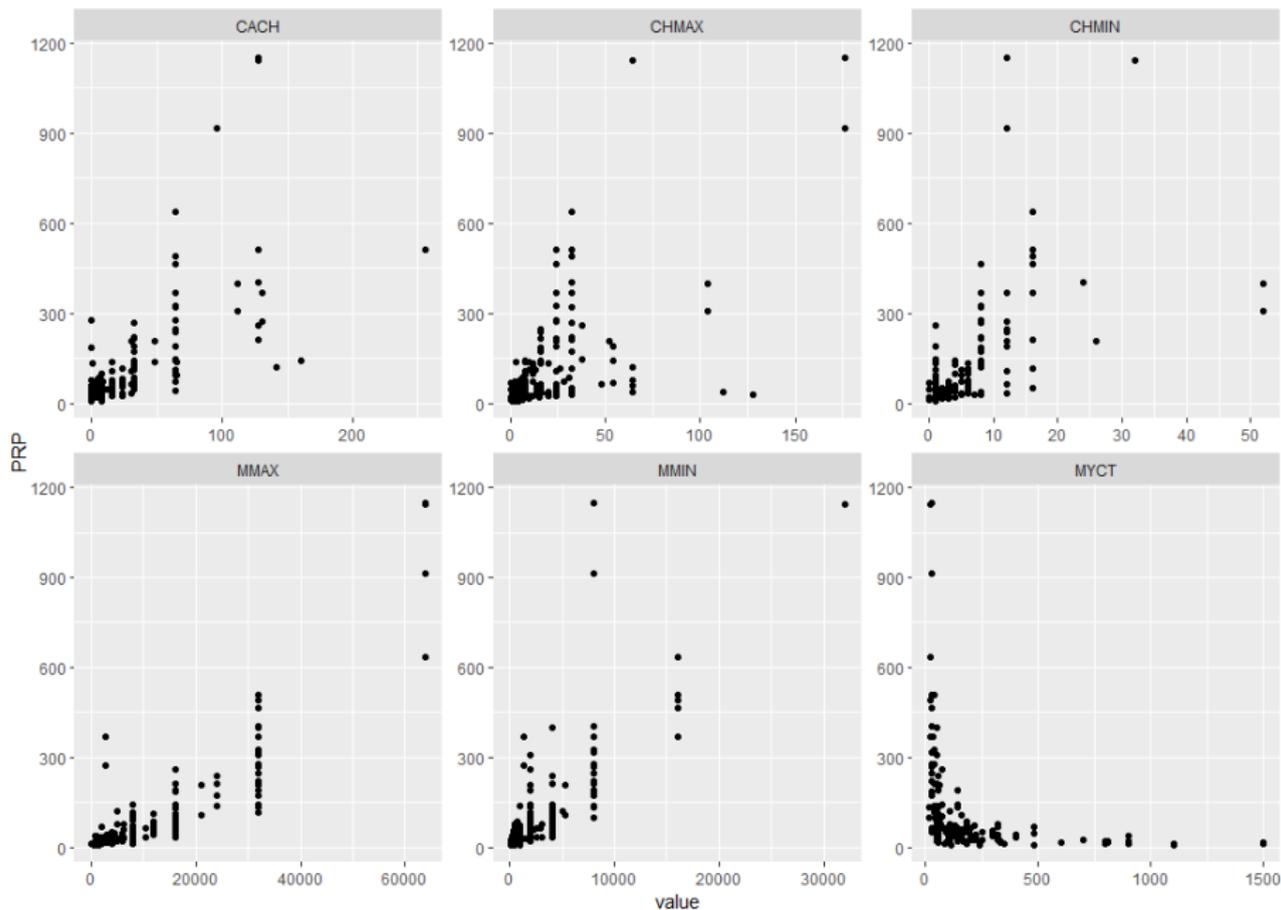
- Comportamento do modelo Gama foi extremamente semelhante ao modelo Normal, evidenciando a robustez do método
- EQM de $\hat{\beta}$ se aproxima de 0 já para tamanhos amostrais baixos
- EQM de $\hat{\gamma}$ se aproxima de 0 para um n a partir de 500, evidenciando sua complexidade

- Dados sobre performance de CPUs(Unidade Central de Processamento) (Ein-Dor and Feldmesser 1987). Variável resposta PRP (performance relativa publicada)
- CACHE(tamanho de memoria cache), CHMIN(número mínimo de canais), CHMAX(número máximo de canais), MCYT(tempo de ciclo da máquina), MMIN(mínimo de memória principal), MMAX(máximo de memória principal).
- 4 modelos testados: Normal, Normal Duplo, Gama Duplo, Normal Inversa Duplo
- Utilização do pacote GAMLSS
- Seleção de variáveis para μ e ϕ via forward stepwise

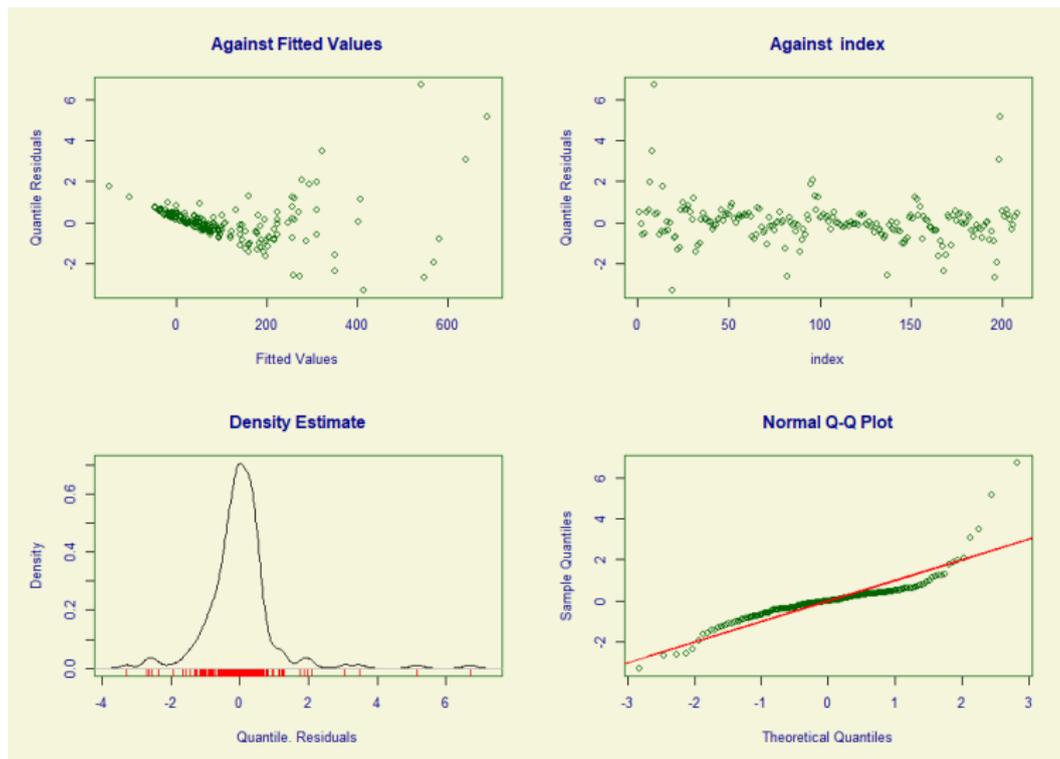
Histograma da variável resposta



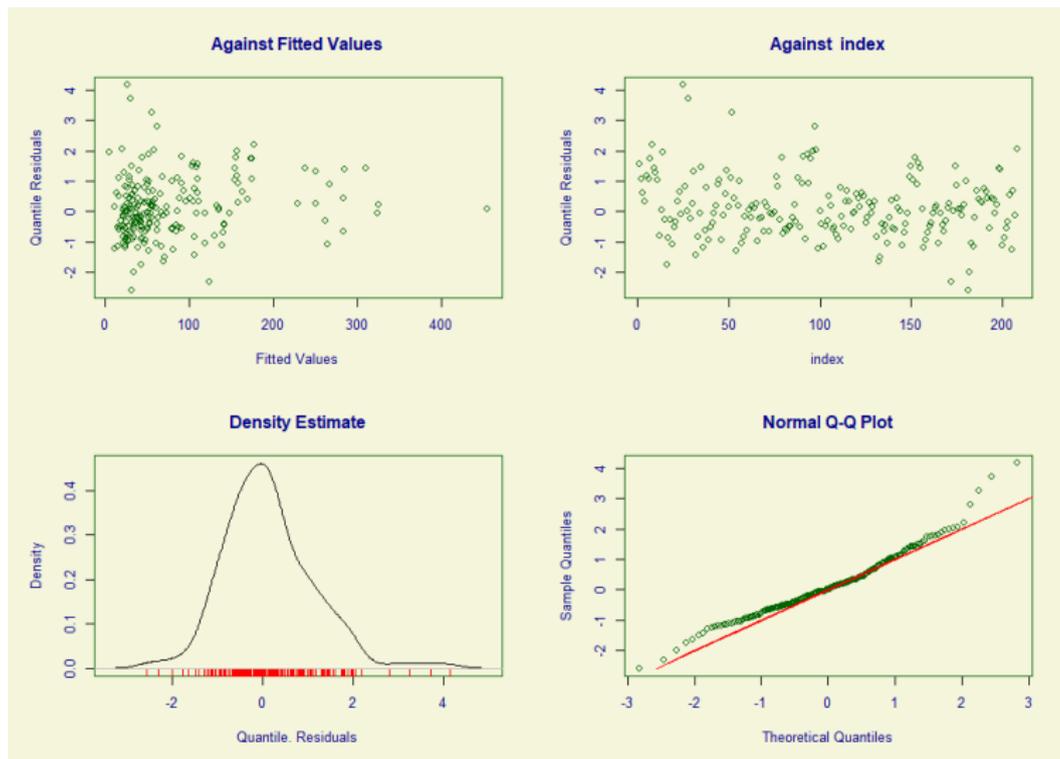
Comportamento em relação as variáveis explicativas



Modelo Normal



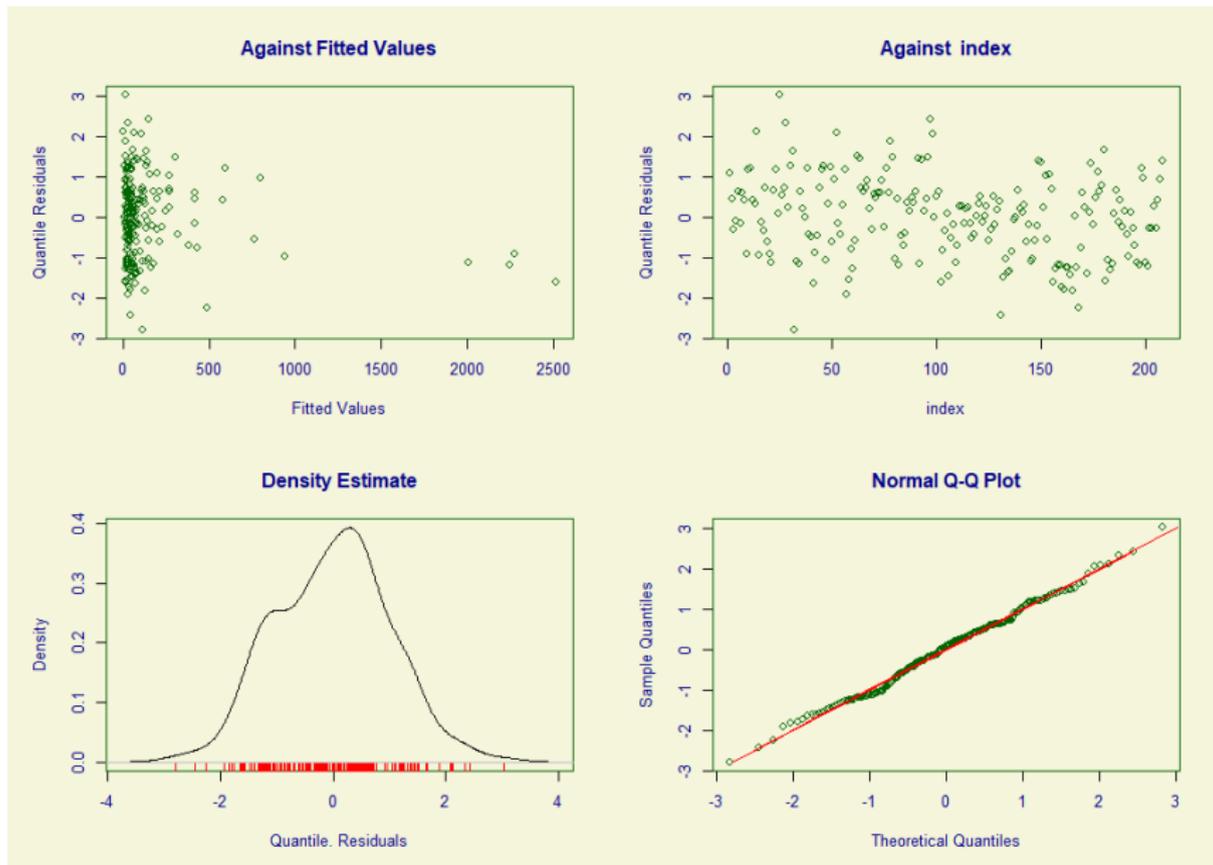
Modelo Normal Duplo



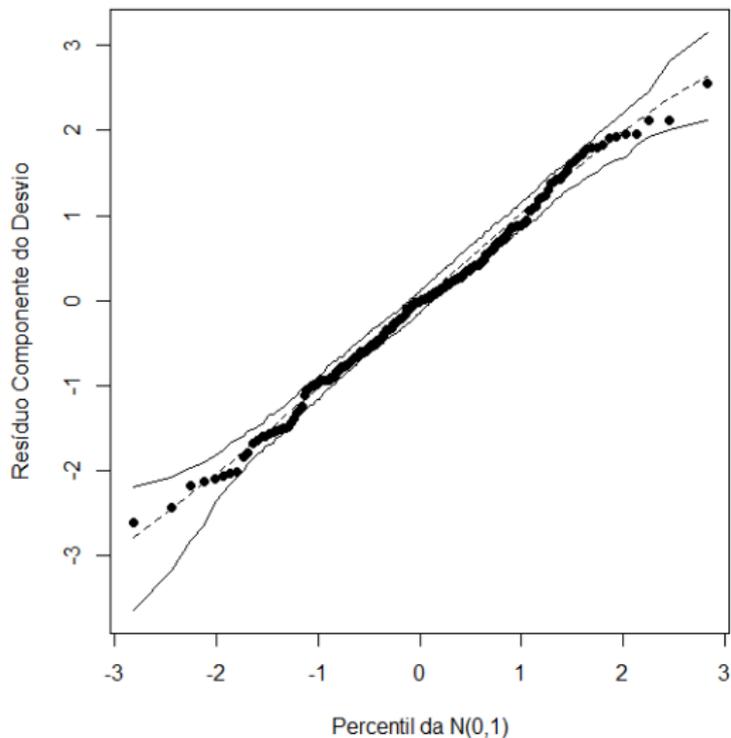
Desempenho dos Modelos Gama e Normal Inversa

Distribuição	Ligação μ	Ligação $\frac{1}{\sqrt{\phi}}$	BIC
Gama	Log	Log	1961.18
Gama	Identidade	Log	-
Gama	Inversa	Log	-
Gama	Log	Identidade	-
Gama	Log	Inversa	1958.77
Gama	Identidade	Identidade	-
Gama	Identidade	Inversa	-
Gama	Inversa	Identidade	-
Gama	Inversa	Inversa	-
Normal Inversa	Log	Log	1966.42
Normal Inversa	Identidade	Log	1968.7
Normal Inversa	Inversa	Log	-
Normal Inversa	Log	Identidade	-
Normal Inversa	Log	Inversa	1929.6
Normal Inversa	Identidade	Identidade	-

O melhor modelo, Normal Inversa, teve o seguinte resultado



Com o seguinte envelope simulado



Coeficientes para $\log(\mu)$

Parâmetro	Estimativa	p-valor
Intercepto	3.543	2e-16 ***
MYCT	-8.888e-04	1.43e-11 **
MMIN	6.651e-05	1.54e-07 ***
MMAX	3.436e-05	2e-16 ***
CACH	8.997e-03	2.38e-12 ***

Coeficientes para $\frac{1}{\sigma}(\sqrt{\phi})$

Parâmetro	Estimativa	p-valor
Intercepto	5.384880219	1.46e-07 ***
MMAX	0.001175228	9.35e-13 **
CHMAX	0.201854466	0.00549 ***

- Ao longo do trabalho, os Modelos Lineares Generalizados Duplos foram profundamente estudados
- Os MLGs duplos foram aplicados em dados simulados, onde vimos o comportamento desses em diferentes tamanhos amostrais
- Os MLGs duplos foram aplicados em dados reais, onde foi possível observar um grande ganho de desempenho em relação aos modelos Gaussianos
- Portanto, os Modelos Lineares Generalizados duplos representam uma ferramenta estatística robusta para modelagem, particularmente quando os conjuntos de dados trabalhados são complexos, heterogêneos e multidimensionais.

- Ein-Dor, Phillip, and Jacob Feldmesser. 1987. "Attributes of the Performance of Central Processing Units: A Relative Performance Prediction Model." *Communications of the ACM* 30 (4): 308–17.
- McCullagh, Peter, and Robert Tibshirani. 1990. "A Simple Method for the Adjustment of Profile Likelihoods." *Journal of the Royal Statistical Society: Series B (Methodological)* 52 (2): 325–44.
- Paula, Gilberto A. 2013. "Modelos de Regressão Com Apoio Computacional. 2013." *Citado Na pág 1* (9): 10.
- Smyth, Gordon K. 1989. "Generalized Linear Models with Varying Dispersion." *Journal of the Royal Statistical Society: Series B (Methodological)* 51 (1): 47–60.
- Smyth, Gordon K, and Arūnas P Verbyla. 1999. "Adjusted Likelihood Methods for Modelling Dispersion in Generalized Linear Models." *Environmetrics: The Official Journal of the International Environmetrics Society* 10 (6): 695–709.